

# HEAT TREATMENT PROCESS PARAMETER ESTIMATION USING HEURISTIC OPTIMIZATION ALGORITHMS

Michael Kommenda<sup>(a,b)</sup>, Bogdan Burlacu<sup>(a,b)</sup>, Reinhard Holecek<sup>(c)</sup>,  
Andreas Gebeshuber<sup>(c)</sup>, Michael Affenzeller<sup>(a,b)</sup>

<sup>(a)</sup> Heuristic and Evolutionary Algorithms Laboratory  
University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria

<sup>(b)</sup> Institute for Formal Models and Verification, Johannes Kepler University,  
Altenberger Straße 69, 4040 Linz, Austria

<sup>(c)</sup> Rübzig GmbH & Co KG, Schafwiesenstrasse 56, 4600 Wels, Austria

[michael.kommenda@fh-hagenberg.at](mailto:michael.kommenda@fh-hagenberg.at), [bogdan.burlacu@fh-hagenberg.at](mailto:bogdan.burlacu@fh-hagenberg.at), [reinhard.holecek@rubig.com](mailto:reinhard.holecek@rubig.com),  
[andreas.gebeshuber@rubig.com](mailto:andreas.gebeshuber@rubig.com), [michael.affenzeller@fh-hagenberg.at](mailto:michael.affenzeller@fh-hagenberg.at)

## ABSTRACT

We present an approach for estimating control parameters of a plasma nitriding process, so that materials with desired product qualities are created. We achieve this by solving the inverse optimization problem of finding the best combination of parameters using a real-vector optimization algorithm, such that multiple regression models evaluated with a concrete parameter combination predict the desired product qualities simultaneously.

The results obtained on real-world data of the nitriding process demonstrate the effectiveness of the presented methodology. Out of various regression and optimization algorithms, the combination of symbolic regression for creating prediction models and covariant matrix adaptation evolution strategies for estimating the process parameters works particularly well. We discuss the influence of the concrete regression algorithm used to create the prediction models on the parameter estimations and the advantages, as well as the limitations and pitfalls of the methodology.

Keywords: parameter estimation, symbolic regression, genetic programming, heuristic optimization

## 1. INTRODUCTION

Nitriding is a case-hardening technique where nitrogen is added to the surface of metal alloys to create a thin hardened nitride layer. The increased surface hardness provides high resistance to wear and crack as well as improved sliding and frictional properties and higher fatigue strength values.

The diffusion of nitrogen into the surface of the metal can take place in the presence of ammonia (NH<sub>3</sub>) using phosphate activation (salt bath and gas nitriding) or in the presence of plasma (highly ionized gas molecules), where nitrogen ions are accelerated to impinge on the metal surface. Plasma nitriding uses pure nitrogen or a mixture of hydrogen and nitrogen that is ionized in intense electric fields, surrounding the surface to be nitriding with ionized gas molecules.

The advantages of the plasma nitriding compared to salt bath or gas nitriding include:

- Lower process temperature, resulting in very small dimensional deformations and distortions
- Higher surface, case and core hardness
- More precise compared to other methods
- Decreased energy demand
- Non-toxic and environmentally friendly, using H<sub>2</sub> and N<sub>2</sub>, with no greenhouse gas emissions

Typically, the treated metals have a predefined set of characteristics that should be reached by the nitriding process:

- Nitriding hardness depth (NHD in mm)
- Surface hardness (SH in HV)
- Thickness of the compound layer (CLT in  $\mu\text{m}$ ).

The results of the nitriding process depend on the material composition and the process control parameters, such as the temperature, process duration, gas concentrations and the parameters of the plasma ionization. These process parameters have to be adjusted precisely to reach the desired material characteristics. The effect of process parameters on the nitriding results is well studied in the case of gas nitriding and allows a close control of the nitriding results (Mittemeijer 1997). However, such a control mechanism for process parameters to achieve the desired results is not available for plasma nitriding.

In this paper, we present a robust and flexible data-driven heuristic optimization approach for finding the optimal control parameters of a plasma nitriding process. In a first step, we create prediction models for each target characteristic based on the available data containing process parameters and material compositions. Then, we use the generated models to solve the inverse problem of finding combinations of process parameters such that the material characteristics after nitriding reach predefined quality values.

Formally, the inverse optimization problem can be defined as follows: given a (regression) model for each of the target characteristics, and a set of quality values to be achieved, is it possible to find the process parameter values so that for each model the desired output is achieved within a specified confidence bound?

This workflow brings a couple of advantages:

1. It exploits the interpolation abilities of the prediction models so that estimates for unknown data can also be given.
2. It allows us to assess the accuracy of the prediction models (in terms of their estimation of product qualities).
3. The results of the process parameter estimation can be further improved by using more accurate prediction models and/or providing more data for training the prediction models.

The whole workflow for process parameter estimation was implemented in HeuristicLab (Wagner 2014), an open-source optimization environment that offers a wide selection of ready-to-use models and algorithms. Our approach for process parameter estimation supports an arbitrary number of models, as well as any configuration of optimizer, optimization targets and optimization bounds.

The remainder of the publication is organized as follows: In the next Section 2 we give a detailed description of the model generation step and discuss the obtained regression models. Section 3 describes in detail the process parameter estimation methodology, Section 4 discusses the obtained results, while Section 5 is dedicated to the conclusion and final remarks.

## 2. REGRESSION MODEL GENERATION

The regression models predicting the three target characteristics NHD, SH, and CLT, which are necessary for solving the inverse optimization problem, were obtained using several data-based modeling algorithms on data acquired from a real-world plasma nitriding process. The dataset contains information about the chemical composition of 14 different working materials, 5 different combinations of process parameters and the achieved plasma nitriding results. In total the dataset contains 70 samples and 18 different measurements for each sample. The samples have been divided into a training and test partition such that the distribution of materials and different process parameters is about the same, resulting in 45 training and 25 test samples.

In our experiments, we used genetic programming (GP) with offspring selection (Affenzeller et al. 2009) and constants optimization (Kommenda et al. 2013) to evolve symbolic regression models. The model length was restricted to 25 tree nodes to improve the interpretability of the generated models. Additional experiments were carried out with a fixed model structure, where the GP algorithm was responsible only for finding the correct variables and constant values.

Further on, the models obtained by GP were compared with other regression models obtained via linear regression (LR) (Seber 2012), support vector machines (SVM) (Cortes 1995), Gaussian processes (GPR) (Rasmussen 2004) and random forest regression (RF) (Breiman 2001) in order to determine each algorithm's performance, generalization ability and suitability for process parameter estimation.

For every regression method we used recommended parameter settings and followed best practices such as grid search with crossvalidation where appropriate and multiple repetitions for each method in order to account for stochastic effects. The specific parameter configurations for the optimization methods enumerated above are described in Table 1.

GP	1000 individuals, maximum tree length 25
LR	No parameters
RF	Grid search intervals: <ul style="list-style-type: none"> <li>- Number of trees <math>n</math> between <math>[2^6, 2^{10}]</math></li> <li>- Feature ratio <math>m</math> in the interval <math>[0.4, 0.7]</math></li> <li>- Sample size ratio <math>r</math> between <math>[0.05, 0.66]</math></li> </ul>
SVM	- NU_SVR regression with an RBF kernel Grid search intervals: <ul style="list-style-type: none"> <li>- Penalty factor <math>C</math> between <math>[2^{-1}, 2^{12}]</math></li> <li>- Kernel function parameter <math>\gamma</math> in <math>[2^{-4}, 2^{-1}]</math></li> <li>- Regularization parameter <math>\nu</math> in <math>[2^{-10}, 1]</math></li> </ul>
GPR	- Constant mean function - Isotropic squared exponential covariance function

**Table 1:** Modeling algorithm parameters

The best models for each modeling algorithm and target characteristic were afterwards selected as follows:

- The best genetic programming models were selected according to the lowest mean absolute error and the lowest model complexity.
- Linear regression creates only one model which was selected.
- The best random forest models were selected out of multiple grid search runs according to the lowest out of bag error estimate.
- The best support vector machine models were selected out of multiple grid search runs according to the lowest cross validation error.
- The best Gaussian process models were selected from multiple runs according to the lowest negative log likelihood.

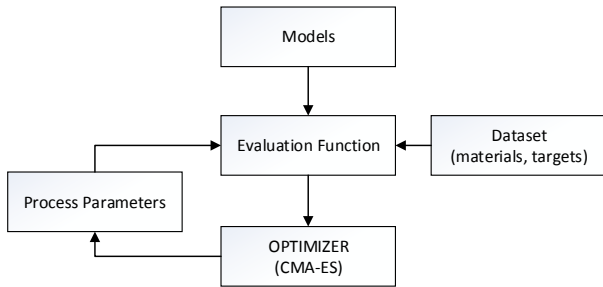
The accuracy of the selected prediction models is stated in Table 2, where for every target characteristic the mean absolute error for the test and training partition is reported. The performance of the different modeling methods varies strongly on the training partition (values inside parenthesis), but with the exception of the linear regression most methods perform similarly on the test partition. This could be explained by the limited amount of data present and we expect that with a larger data basis the modeling accuracies will increase.

	NHD	SH	CLT
GP	0.030 (0.026)	47.65 (51.37)	1.296 (0.772)
LR	0.042 (0.033)	54.42 (42.68)	1.865 (1.140)
RF	0.039 (0.015)	53.36 (17.03)	1.176 (0.627)
SVM	0.031 (0.011)	54.83 (8.370)	1.101 (0.518)
GPR	0.029 (0.014)	52.74 (12.46)	1.208 (0.363)

**Table 2:** Test and training (inside parenthesis) mean absolute error of the obtained models for each of the target characteristics.

### 3. PROCESS PARAMETER ESTIMATION

Process parameter estimation is made possible through the interaction of the generated regression models from the previous section with several other components including the dataset, the evaluation function and the optimizer that performs the actual search for real-valued parameter combinations.



**Figure 1:** Process Parameter Optimization Flowchart

Figure 1 illustrates the interplay between the individual components when process parameter estimation is performed. The optimization algorithm generates multiple combinations of process parameters within predefined bounds for each parameter. Every parameter combination is evaluated by the aforementioned prediction models including data about the processed material. The deviations between the predicted and desired qualities are passed back to the optimization algorithm, which uses that information to generate new parameter combinations. The optimization algorithm minimizes the deviations by the iterative generation of new process parameter combinations. The procedure is outlined in pseudocode in Algorithm 1.

This general workflow is independent of the concrete data-based modeling algorithm for generating the prediction and the optimization algorithm for tuning the process parameters to reach the desired product qualities. However, we observed that some algorithms are better suited than others for generating the prediction models.

---

#### Algorithm 1: Process parameter optimization workflow

---

**Input:** material compositions, desired target values  $T$ , prediction models  $M$

**Output:** Optimal parameter combination  $\vec{p}$

```

1  $\vec{p} \leftarrow$  empty parameter vector;
2 while optimizer not converged do
3   generate new real-valued parameter vector  $\vec{v}$ ;
4   use models  $M$  to get estimated values  $E$  with
   parameters  $\vec{v}$  and given material compositions;
5   evaluate objective function  $f(E, T)$ ;
6   if  $f(E, T) <$  best value so far then
7      $\vec{p} \leftarrow \vec{v}$ 
8 return  $\vec{p}$ ;
  
```

---

For example, random forest regression generates highly accurate prediction models, but those models are difficult to use for process parameter optimization as their response can contain instabilities, which makes it harder for the optimization algorithm to find good process parameter settings. In contrast to random forest regression, symbolic regression by genetic programming produces more stable prediction models, especially when the complexity of the models is rather low, and hence it is easier for the optimization algorithm to generate appropriate process parameter settings.

In our experiments three target characteristics (NHD, SH, and CLT) should be obtained by the nitriding process through the estimation of five process parameters: the process duration, temperature, average plasma power, relative amount of nitrogen, and the applied pressure. We used the covariant matrix adaptation evolution strategy – CMA-ES (Hansen 2003) as a real-valued optimization algorithm for generating the estimates for the process parameters.

The objective function to be minimized by the CMA-ES was defined such that each deviation between the desired target value  $T_i$  and estimated  $E_i$  had an equal influence on the final objective value. This was achieved by scaling the values by the variance of the target characteristic  $Var_{T_i}$  over all samples and aggregating the deviations into a single error measure:

$$f(E, T) = \sum_{i=1}^N \frac{(T_i - E_i)^2}{Var_{T_i}}$$

**Equation 1:** Aggregated error measure

As shown in Figure 1, the evaluation function is arguably the most important component in our optimization approach. In this publication, the aggregated error measure  $f(E, T)$  will be used to evaluate the models and investigate their influence on the process parameter estimation.

#### 4. RESULTS

In this section we analyze the influence of the prediction models on the accuracy of the process parameter estimation. We kept the CMA-ES real-valued optimizer (although using other real-valued optimization algorithms is possible as well) for generating the process parameter estimations while exchanging the prediction models for evaluating the process parameters according to the objective function discussed in the previous section and given by Equation 1.

The CMA-ES was set to run for a maximum of 200 generations with a population of 20 individuals, yielding 4000 different process parameter combinations. The algorithm converged in all cases within a third of a second. Although the CMA-ES was able to find parameter combinations that minimized the target for each of the used models, variations due to the scaling of the model targets and the aggregated error measure were present.

The accuracy of the estimated process parameters was evaluated on the test partition of the dataset, which has not been used for model training nor selection. A number of 20 repetitions of the CMA-ES algorithm have been performed to account for the stochasticity of the algorithm and the presence of multiple local optima. Multiple parameter combinations achieve target characteristics equally well, especially when estimating several process parameters simultaneously. Table 3 shows the average absolute deviation of the estimated parameter to the original ones on the training partition and Table 4 shows the same for the test partition.

	GP	LR	RF	SVM	GPR
Duration	5.45	3.75	5.75	2.60	2.66
Temp.	41.72	32.18	31.57	23.14	26.47
Plasma power	16.53	15.52	25.14	16.06	9.69
N <sub>2</sub>	0.15	0.16	0.12	0.11	0.11
Pressure	0.52	0.41	0.53	0.26	0.24

**Table 3:** Average mean absolute error of the predicted process parameters on the training data.

	GP	LR	RF	SVM	GPR
Duration	4.65	2.91	6.53	4.60	4.18
Temp.	42.05	42.18	35.42	33.55	36.75
Plasma power	22.14	15.64	24.96	13.50	14.43
N <sub>2</sub>	0.19	0.16	0.16	0.17	0.22
Pressure	0.27	0.26	0.62	0.32	0.27

**Table 4:** Average mean absolute error of the predicted process parameters on the test data.

With the exception of GP and LR, the models perform significantly better on the training partition compared to the test partition. The difference between the training and test qualities of regression models means that some models are slightly overfit on the training set and may exhibit poor interpolation capabilities on new data. Therefore, confidence margins must also be taken into account when evaluating the models, along with their error on the test set.

Confidence margins were calculated by counting how many times the estimated process parameter values were within 10% of the target value. To account for the different value ranges of each process parameter, the deviation was scaled by the range of the parameter and a hit was reported if the scaled deviation is below 10%. Tables 5 and 6 show the 10% hits for each type of models calculated on the training and test partitions.

	GP	LR	RF	SVM	GPR
Duration	28	25	46	34	32
Temp.	26	21	46	30	24
Plasma power	34	36	41	32	40
N <sub>2</sub>	36	23	43	32	30
Pressure	20	25	38	35	34

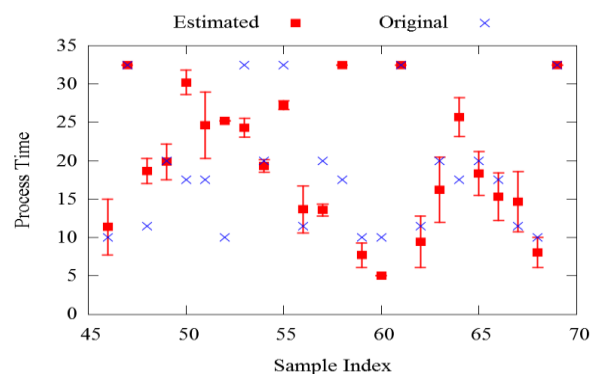
**Table 5:** 10% confidence hits on the training data.

	GP	LR	RF	SVM	GPR
Duration	15	15	23	12	13
Temp.	14	9	23	6	10
Plasma power	16	17	19	19	17
N <sub>2</sub>	17	11	18	12	9
Pressure	14	17	17	11	16

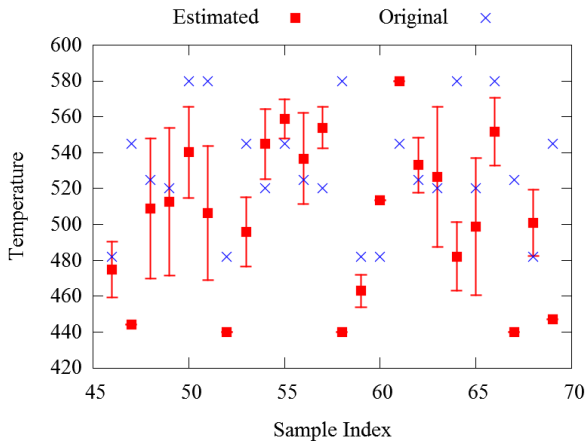
**Table 6:** 10% confidence hits on the test data.

As duration and temperature have the greatest influence on the nitriding outcome, their errors and confidence margins constitute the main criteria for choosing the most suitable models. From this perspective, GP and RF distinguish themselves and will be considered further.

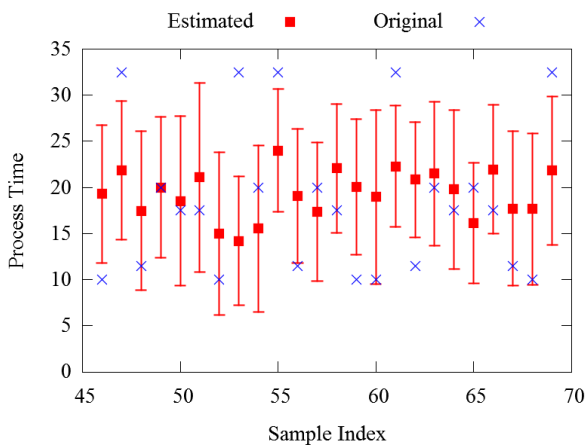
We notice that on the test data RF produces the most confidence hits for both temperature and process duration, with GP on the second place. This result is mainly explained by the large variance of the RF estimation results. The average and standard deviations for the estimations of the process duration and temperature is visible Figures 2 and 3 for GP and in Figure 4 and 5 for RF respectively, where the amplitude of the candlesticks represents the standard deviation of the estimations at the given data row. Therefore, the models produced by GP are more stable and more suitable for the optimization of process parameters.



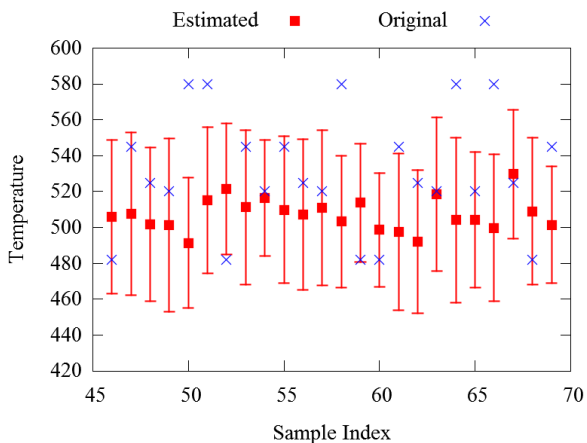
**Figure 2:** Process duration estimations using GP models on the test data.



**Figure 3:** Temperature estimations using GP models on the test data.



**Figure 4:** Process duration estimations using RF models on the test data.

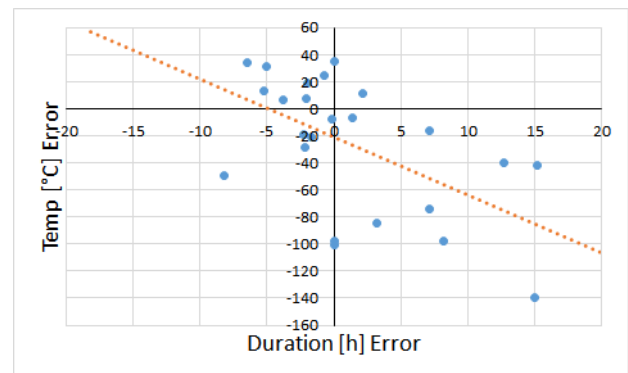


**Figure 5:** Temperature estimations using RF models on the test data.

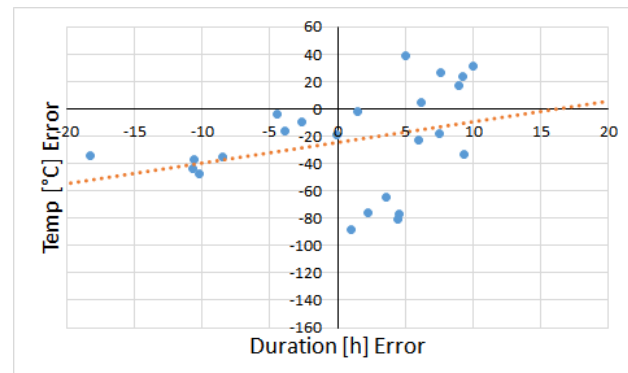
Furthermore, the estimation quality of process parameters depends on the number of parameters allowed as inputs to the models. As interdependencies exist between the parameters, more parameters will determine an increased number of local optima and consequently more variance of their values. We illustrate one such interdependency between two antagonistic process parameters:

duration and temperature. From the physical standpoint, a higher temperature allows the nitriding process to complete in less time. Conversely, a lower temperature means that the process will require more time to achieve the desired results.

This relationship should be reflected in the results of our process parameter estimation. To illustrate this, we used the error between the original and the estimated values for duration and temperature. Using the GP models, we notice in Figure 6 that when one of the parameters has a lower value – therefore a negative normal error with respect to the target value, the optimizer compensates by increasing the value of the other. This means that although the estimated values are not entirely accurate on the test data, the optimizer together with the GP models manage to obtain physically valid results, correctly reflecting the inverse physical relationship between duration and temperature. This relationship was not so easily noticeable for the models generated by RF and shown in Figure 7. Therefore, GP models are more suitable for process parameter estimations from both a pure data-mining perspective and from a physical modeling perspective.



**Figure 6:** Duration-temperature relationship estimated using GP models



**Figure 7:** Duration-temperature relationship estimated using RF models

## 5. CONCLUSION

In plasma nitriding, the goal is to obtain desired material characteristics such as surface hardness, nitriding hardness depth or thickness of the compound layer by tuning the process parameters for each treated material.

In this paper, a two-step optimization approach was presented for finding the optimal control parameters of a plasma nitriding process. In a first step, regression models were generated for each of the target characteristics using a number of different data-based modeling algorithms on data acquired from the process. These models were then used to solve the inverse optimization problem of finding optimal process parameter combinations that would lead to predefined target values for the material characteristics.

The second step of the methodology involved a separate optimization algorithm used to minimize an objective function that aggregated each model's predictions into a unified error measure. To ensure that each model contributes equally to the objective function, for each model the predicted values were scaled according to the variance of the target characteristic.

We performed several experiments to validate the methodology, where we tested GP regression models against models obtained using linear regression, support vector machines, random forests and Gaussian processes. Our models and objective function were incorporated into a CMA-ES optimizer which was used to estimate process parameters on a separate test dataset.

Experimental results have shown that the variance of the estimated values increases with the number of process parameters used as inputs for the regression models. This fact is explained by the increased number of local optima which the optimizer has to overcome when larger combinations of process parameters are estimated. In our experiments we used combinations of five process parameters with the CMA-ES optimizer.

The results concerning two of the most important process parameters, duration and temperature, confirm the fact that GP models with a simpler structure were better suited for solving the inverse optimization problem. Although on average, other models such as RF, SVM and GPR produced better results on the training data, these models proved less effective for the estimation of process parameters on the test data, as shown by the large variance of the estimated values in the case of RF, or the lower number of estimations within the 10% confidence interval in the case of LR, SVM and GPR.

Furthermore, the feasibility of the estimated process parameter combinations was analyzed from a physical perspective. The generated GP models correctly reflected the inverse physical relationship between duration and temperature, which was not as obvious in the case of the other models.

In conclusion, our process parameter optimization approach works reasonably well on real-world data acquired from the chemical process itself. Our optimization setup is flexible as it allows any type of regression model or even regression ensembles to be used for the parameter estimation part, in conjunction with any real-vector

optimizer such as the CMA-ES. Future research directions involve further development of this approach by testing other optimizers and other combinations of regression models.

## ACKNOWLEDGMENTS

The work described in this paper was done within DAMOPLAS, a research project financed by the Austrian Research Promotion Agency (FFG Basisprogramm).

## REFERENCES

- Affenzeller, M., Winkler, S., Wagner, S., and Beham, A. (2009). *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*. Chapman & Hall/CRC. ISBN 978-1584886297.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45 (1): pp. 5-32.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*. 20: 273-297
- Hansen, Nikolaus, Müller, Sibylle D., and Kourmoutsakos, Petros (2003): Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), *Evolutionary Computation*, 11(1), pp. 1-18.
- Kommenda, M., Kronberger, G., Winkler, S., Affenzeller, M., and Wagner, S (2013) Effects of constant optimization by nonlinear least squares minimization in symbolic regression. *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*, pp. 1121-1128. ACM, 2013.
- Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. University of Michigan Press.
- Mittemeijer, E. J., & Somers, M. A. (1997). Thermodynamics, kinetics, and process control of nitriding. *Surface Engineering*, 13(6), pp. 483-497.
- Rasmussen, Carl Edward. "Gaussian processes in machine learning." *Advanced lectures on machine learning*. Springer Berlin Heidelberg, 2004. 63-71.
- Seber, George AF, and Alan J. Lee. "Linear regression analysis". Vol. 936. John Wiley & Sons, 2012.
- Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., and Affenzeller, M. (2014) Architecture and Design of the HeuristicLab Optimization Environment. In *Advanced Methods and Applications in Computational Intelligence, Topics in Intelligent Engineering and Informatics Series*, pp. 197-261. Springer